

DOCUMENT RESUME

ED 170 337

TH 008 650

AUTHOR Pagan, Barbara M.; Horst, Donald P.
TITLE Selecting a Norm-Referenced Test. ESEA Title I
Evaluation and Reporting System. Technical Paper No.
4.
INSTITUTION RMC Research Corp., Mountain View, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Oct 76
NOTE 6p.
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Evaluation Criteria; National
Norms; *Norm Referenced Tests; Testing; Test
Reliability; *Test Selection; Test Validity
IDENTIFIERS Content Validity

ABSTRACT

Brief guidelines for the selection of norm referenced tests are offered. Four considerations in selecting an appropriate test are: (1) appropriateness of test content; (2) appropriateness of the norm group; (3) test reliability; and (4) ease of administration.
(MH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SELECTING A NORM-REFERENCED TEST

FSEA Title I Evaluation and Reporting System

Technical Paper No. 4

**Barbara M. Fagan
Donald P. Horst**

October 1976

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

RMC Research Corporation
Mountain View, California

Norm-referenced tests (NRT) are commonly used in evaluating educational treatments because these tests provide the evaluator with a kind of quasi-control or comparison group. An NRT is one that has been given to a group of individuals (or a representative sample from the group) in order to establish test score norms for the group. The norms reflect the distribution of scores made by the individuals in the norming sample on each part of the test.

Any test for which normative data have been collected can be considered an NRT. Usually, however, the term refers to commercially published achievement tests. The major publishers collect test scores from nationally representative samples of students and provide the test user with tables of means, standard deviations, and other properties of the norm-group score distributions. Alternatively, a State, school district, or other agency may collect test scores from local students and calculate the norms for this group. In any case, the scores of the norm group provide standards against which various treatment groups can be compared.

In selecting an NRT, the user should consider:

Content of the test

Appropriateness of the norming sample

Reliability

Ease of administration

These four points are discussed below.

Test Content

The content of the test should be consistent with the purpose of the evaluation. The major commercial tests provide good general measures of the skills they were designed to test. However, a particular test may not cover some of

the specific skills or objectives emphasized in your program. Locally developed or special purpose tests may have the opposite problem. That is, they may measure specific project objectives very well but fail to give a comprehensive overview of achievement in the subject area.

There is no fixed procedure for determining whether test content is appropriate for your purposes. At a minimum, teachers or others familiar with the test should review the items carefully before selecting any test.

Norming Sample

In norm-referenced evaluation, the treatment group is compared to the norm group. It follows that the norm group should be one with which comparisons are meaningful. Norm groups for the major standardized achievement tests represent the entire national population of students. It is almost always interesting to know how well your students are doing in relation to national norms, but if your treatment group students are very different from the students who make up the bulk of the norm group, care should be taken in interpreting results.

Some subgroups of students may systematically improve in relation to the rest of the norm group. Other subgroups may lose ground. Of course, if your treatment group were like those which lose ground compared to the norms, then a project which permitted them to hold their relative position would be an effective educational effort. If evaluated using the Title I norm-referenced evaluation model, however, the project would appear to be unsuccessful.

This type of situation illustrates an important concept in norm-group comparisons. It may be of interest to compare the progress of your

students to that of the national population. But if a norm group is to provide a realistic no-treatment expectation for your students, the students composing it must be very much like your students and be in an educational program that is very similar to yours except for the special Title I treatment. Commercially published tests cannot provide norm groups to match every situation, and the development of local norms may be advisable.

Reliability

Reliability is the extent to which a test measures the same thing each time it is used. All major standardized tests have been carefully developed to maximize reliability. Only when locally developed or specialized tests are used should reliability be a major concern in selecting a test.

Where no publisher's data on reliability are available, a judgment by an experienced test constructor should be obtained. A simple item analysis and a check on reliability (e.g., split-half method) can then be included in the pretest data analysis.

Ease of Administration

Most commercial NRTs are relatively easy to administer. However, before selecting a test, the format for administering the test (group or individual), the types of answer sheets, the length of time required, and any other special requirements should be considered. The number of different levels required to test a given group affects logistics and scheduling. At least one test publisher provides many different levels, but organizes the instructions and timing so that several levels may be given at the same time in a single classroom.